Bayesian Hierarchical Modelling of Young Stellar Clusters

Javier Olivares Romero

Université Grenoble Alpes & Universidad Nacional de Educación a Distancia

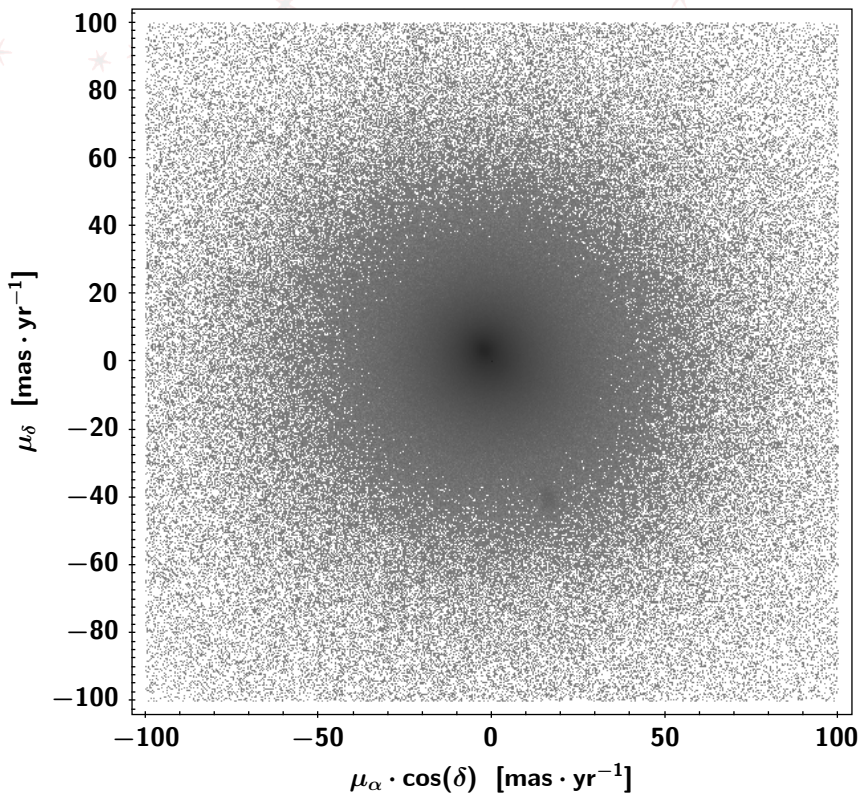
Supervisors:
Estelle Moraux & Luis Manuel Sarro Baro

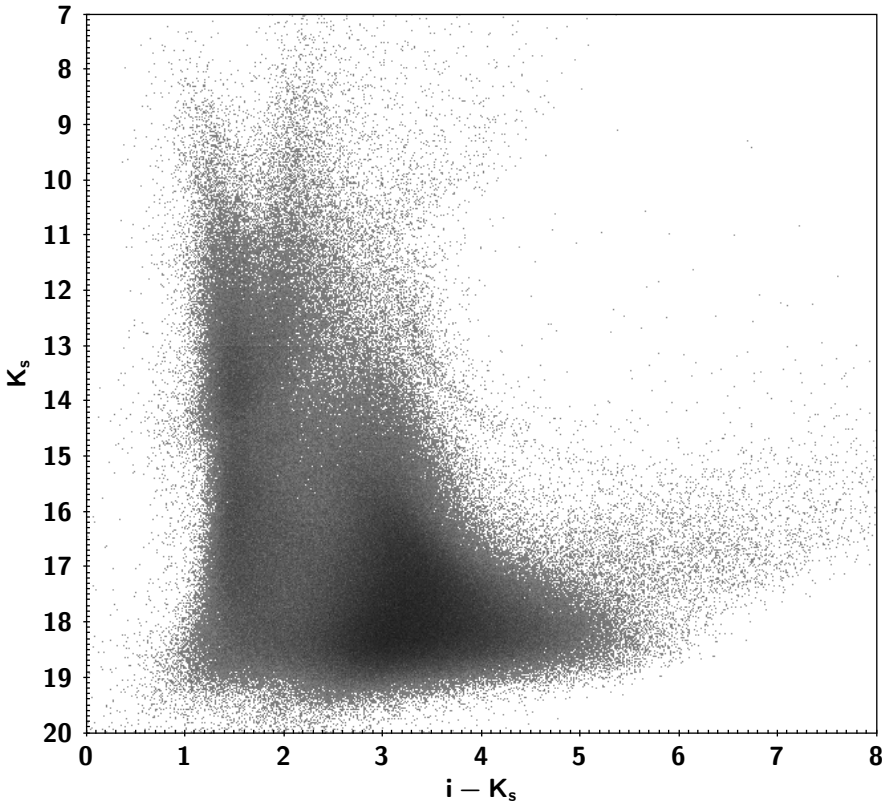
October 18, 2017

Stellar clusters: Origins

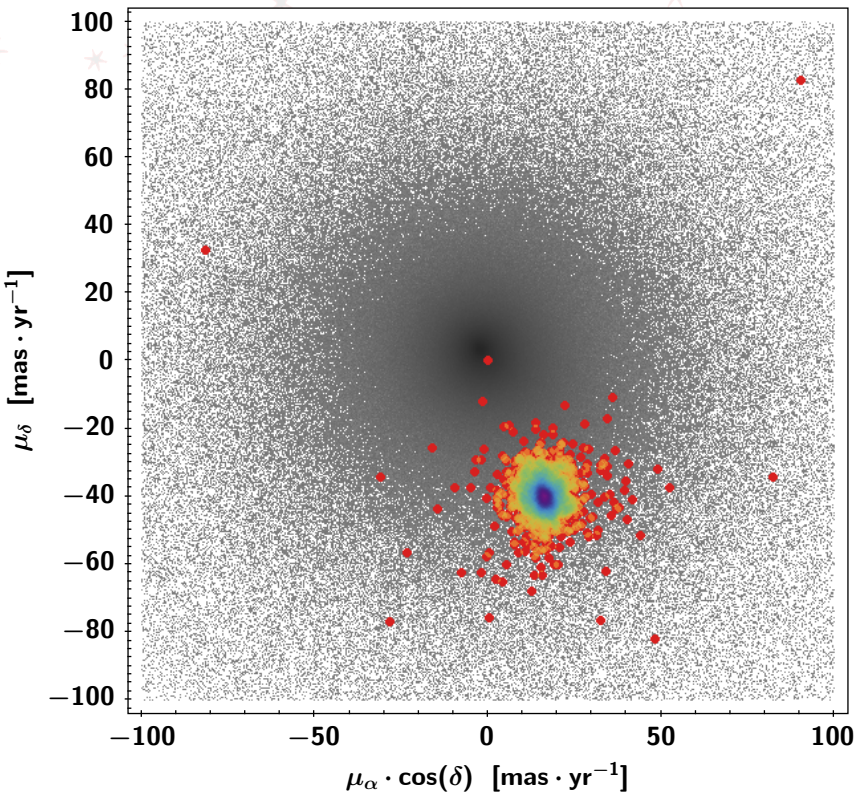
Stellar clusters: Origins

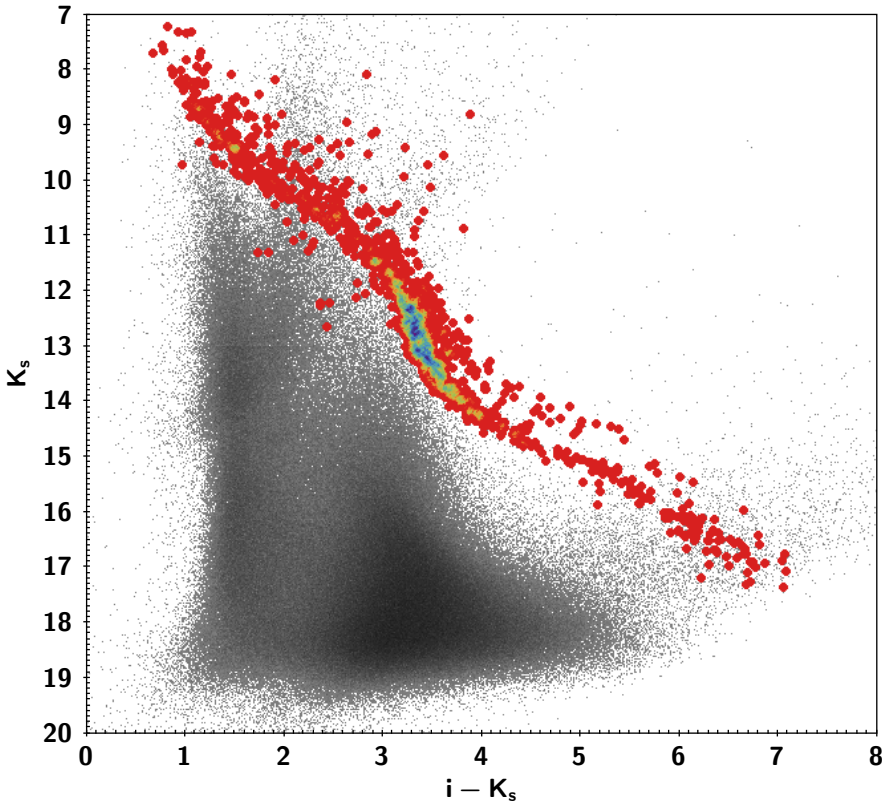
Stellar clusters: Observables





Stellar clusters: Observables





Objective

Obtain the statistical distributions of the cluster population

- Sky positions
- Proper Motions
- Apparent magnitudes

As a by-product:

- Membership probabilities

Current approach

- BANYAN I (Malo et al. 2013)
- UPMASK (Krone-Martins & Moitinho 2014)
- Sarro (Sarro et al. 2014)
- BANYAN II (Gagné et al. 2015)
- Sampedro (Sampedro & Alfaro 2016)
- LACEwING (Riedel et al. 2017)

Caveats

- Classifiers only
- Homogeneous uncertainties
- Fully observed objects

Current approach

- BANYAN I (Malo et al. 2013)
- UPMASK (Krone-Martins & Moitinho 2014)
- Sarro (Sarro et al. 2014)
- BANYAN II (Gagné et al. 2015)
- Sampedro (Sampedro & Alfaro 2016)
- LACEwING (Riedel et al. 2017)

Caveats

- Classifiers only
- Homogeneous uncertainties
- Fully observed objects

Remedy

- Cluster distributions
- Heterogeneous uncertainties
- Missing value objects

Strategy

1. Create a new method.
2. Benchmark it.
3. Validate it
4. Evaluate it.
5. Discussion of the results.

- Bayesian Hierarchical Model
- Pleiades cluster
- Synthetic data
- DANCe DR2

Pleiades DANCe DR2 data set

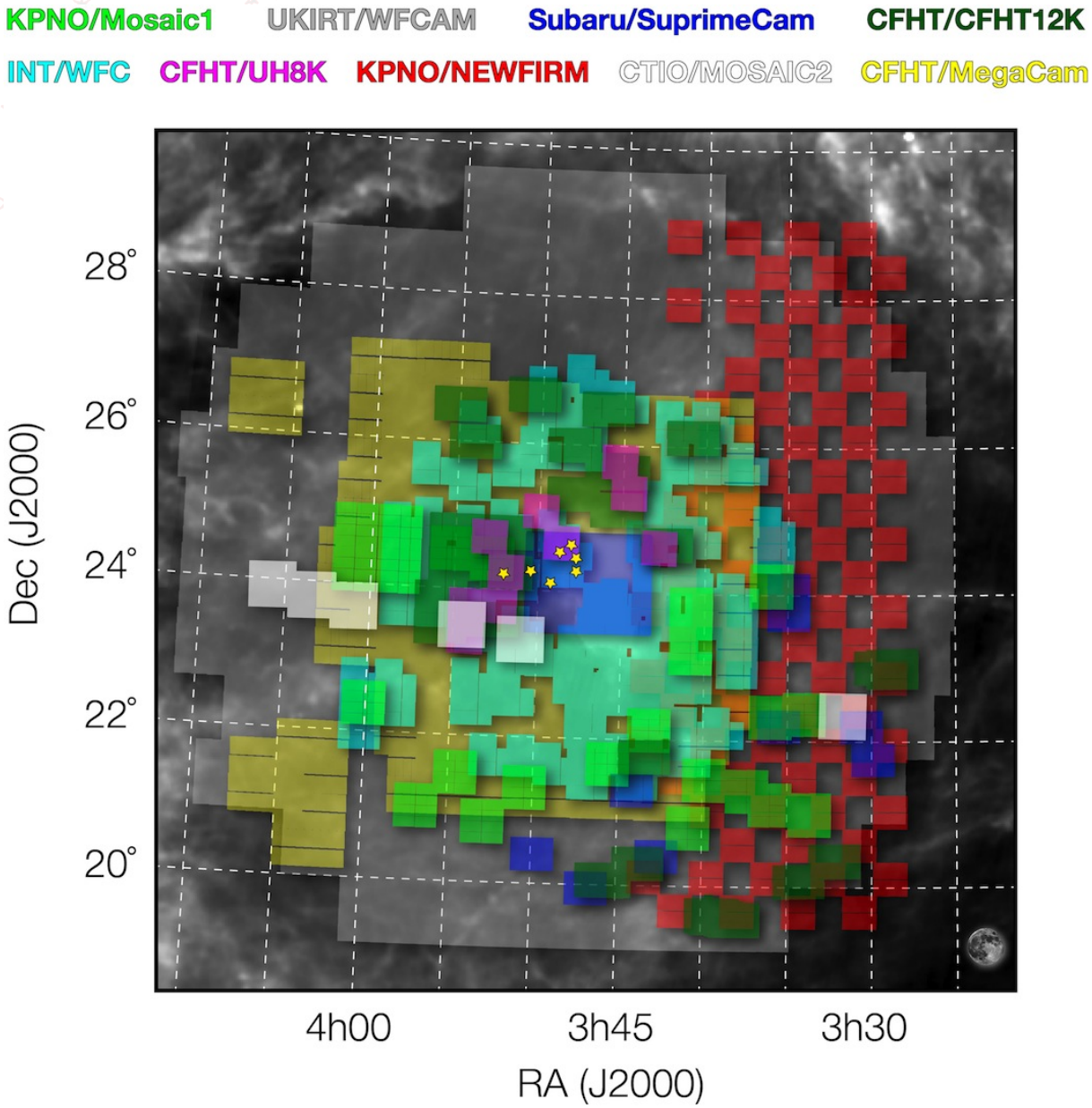
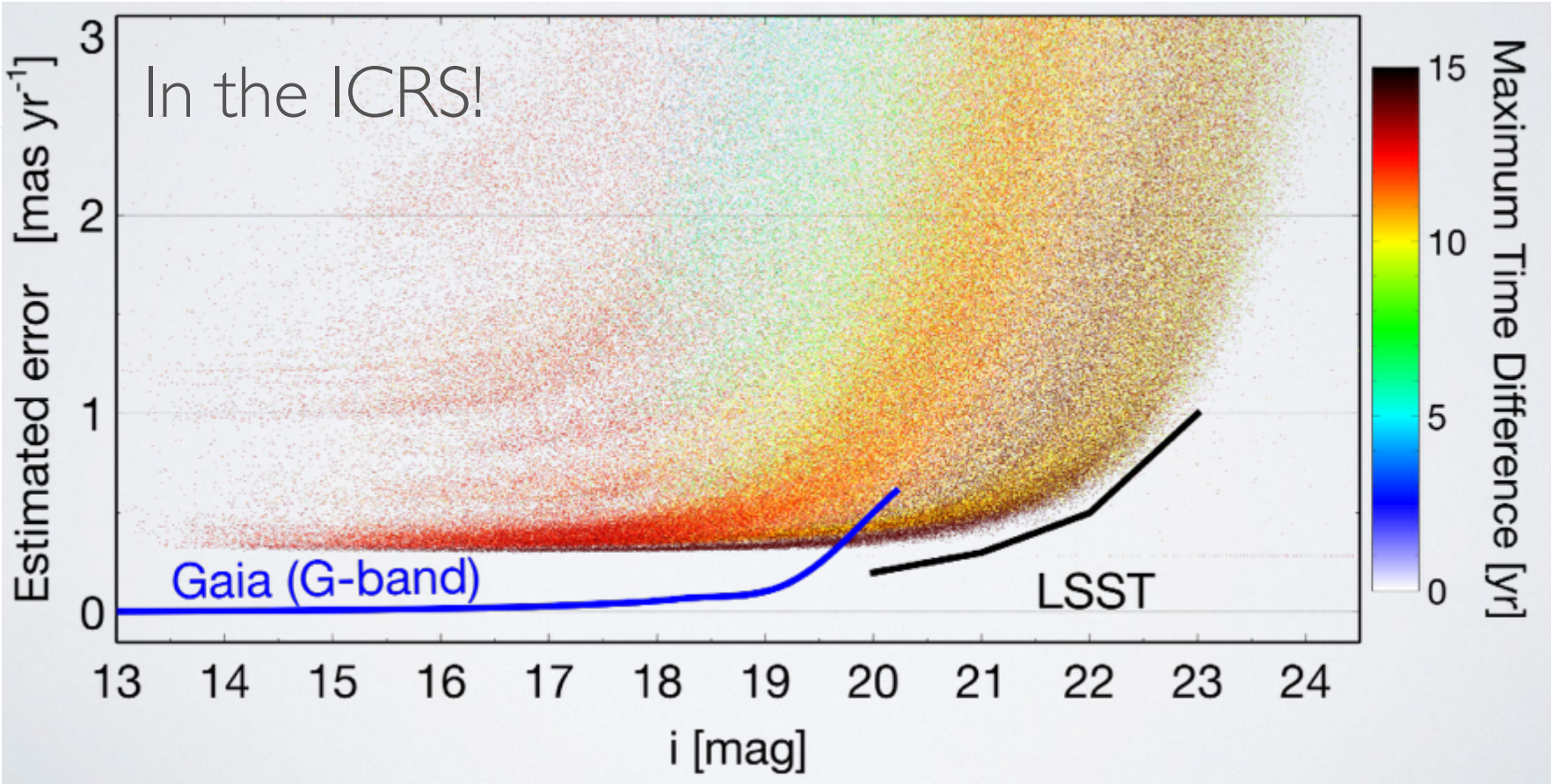


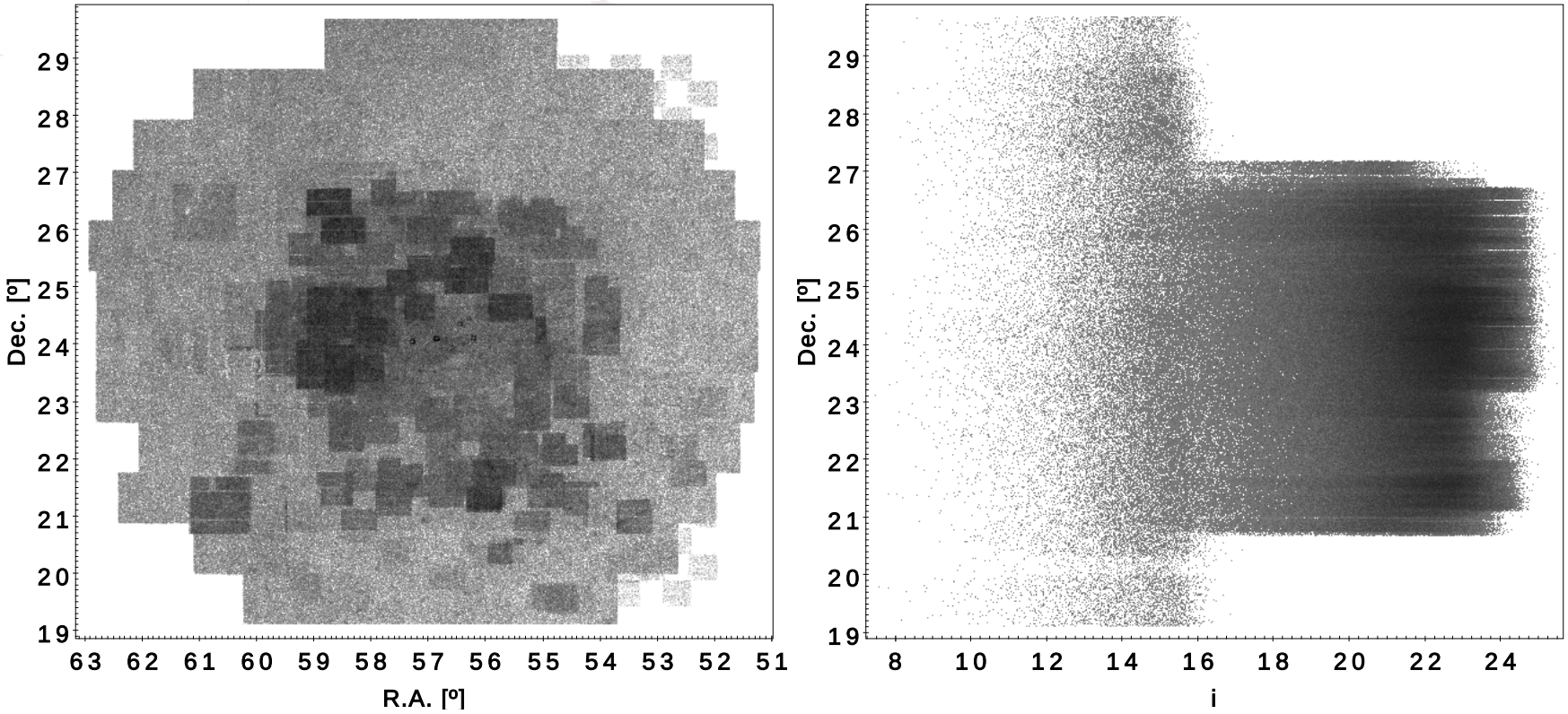
Image by Bouy et al. 2013

Pleiades DANCe DR2 data set



Credit: Hervé Bouy

Pleiades DANCe DR2 data set



Restricted data set

Selection

- Membership probabilities of Bouy et al. 2015
- Size: 10^5 objects.
- Probability threshold: 1×10^{-11}

Assumptions

- No cluster member outside the sample
- Completeness limits of DANCe DR2 apply to the cluster

Overview of the method

Bayes' theorem

$$p(\boldsymbol{\theta}|\mathbf{D}, \mathcal{M}) = \frac{p(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{D}|\mathcal{M})}$$

Hierarchical model

Generative model

True values + Heteroscedastic data

Missing values

Types

- Deterministic
- Stochastic

Simplistic assumptions

- Ignorability
- Missing at random

Details of the method

Selection of observables

- Proper motions and $i - Ks, Y, J, H, K_s$
- Sky positions (independent analysis)

Generative model

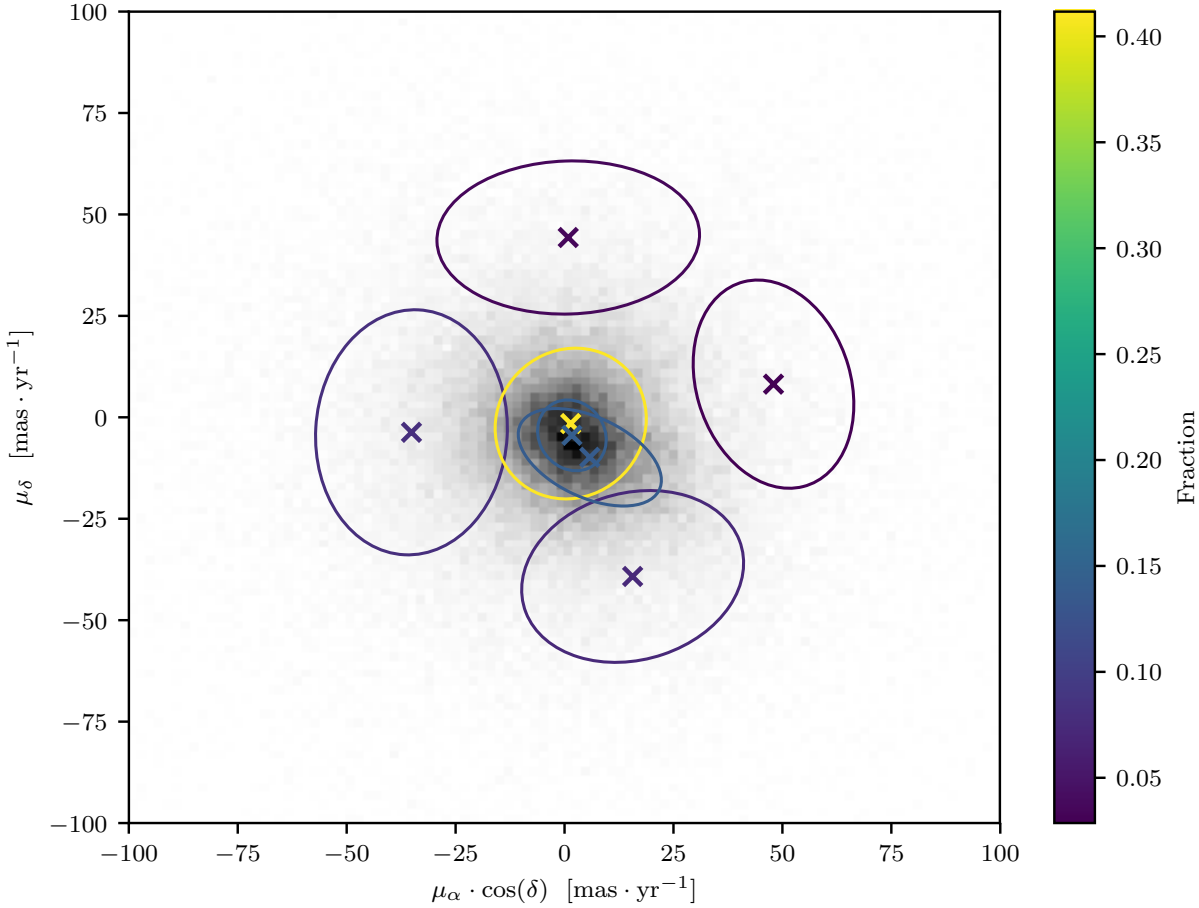
Field + Cluster

Assumptions

- Observables of Sarro et al. 2014 are correct.
- Held the field fixed.
- Photometry independent of proper motions.
- Populations: Single stars and Equal-Mass Binaries (EMB).

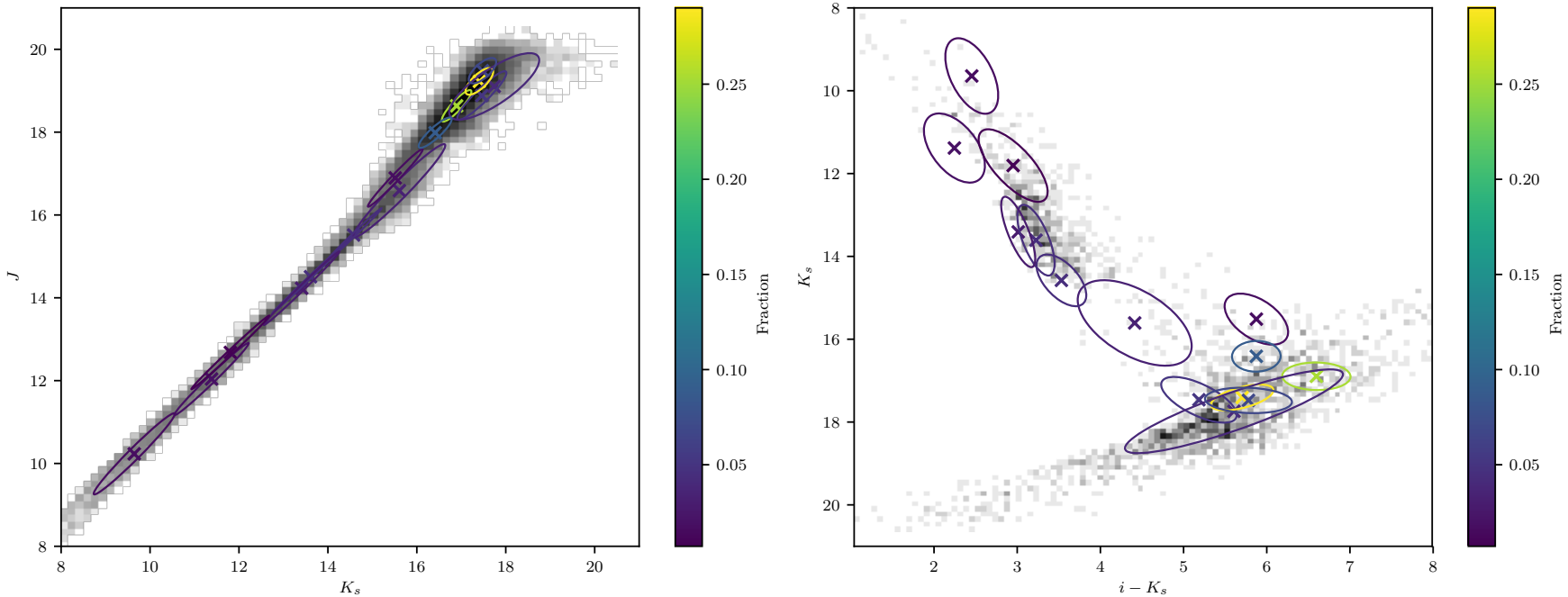
Field: Proper Motions

Gaussian Mixture Model + Uniform with 8 components (BIC)
Maximum-Likelihood Estimate



Field: Photometry

Missing values Gaussian Mixture Model with 14 components (BIC)
Maximum-Likelihood Estimate



Cluster model

Single stars + Equal-Mass binaries

Proper motions:
GMM

- Singles: 4 components
- EMB: 2 components

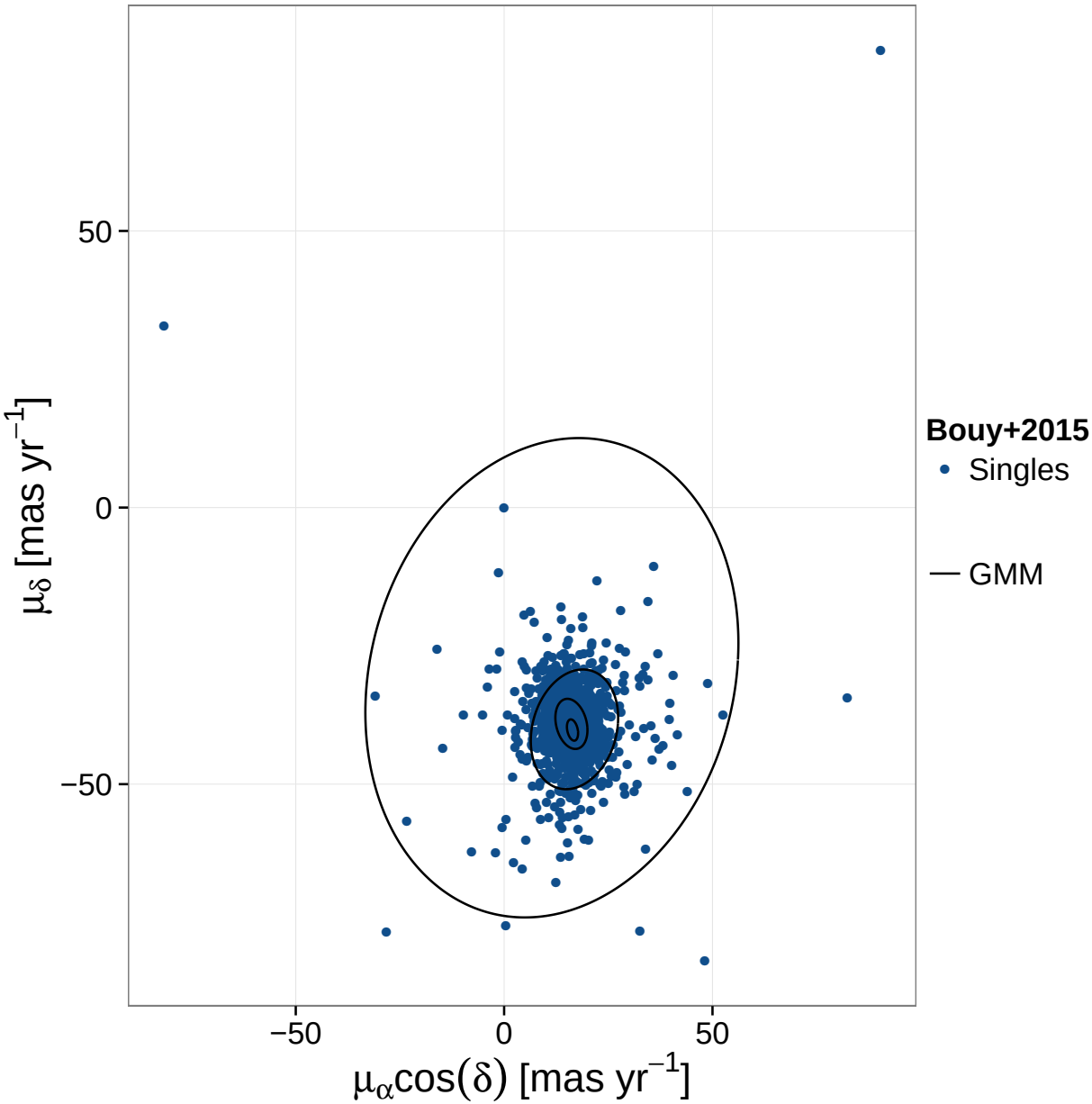
Photometry:
5D Multivariate Gaussian

- True $i - K_s$ distribution
- Splines; Y, J, H, K_s
- Intrinsic dispersion

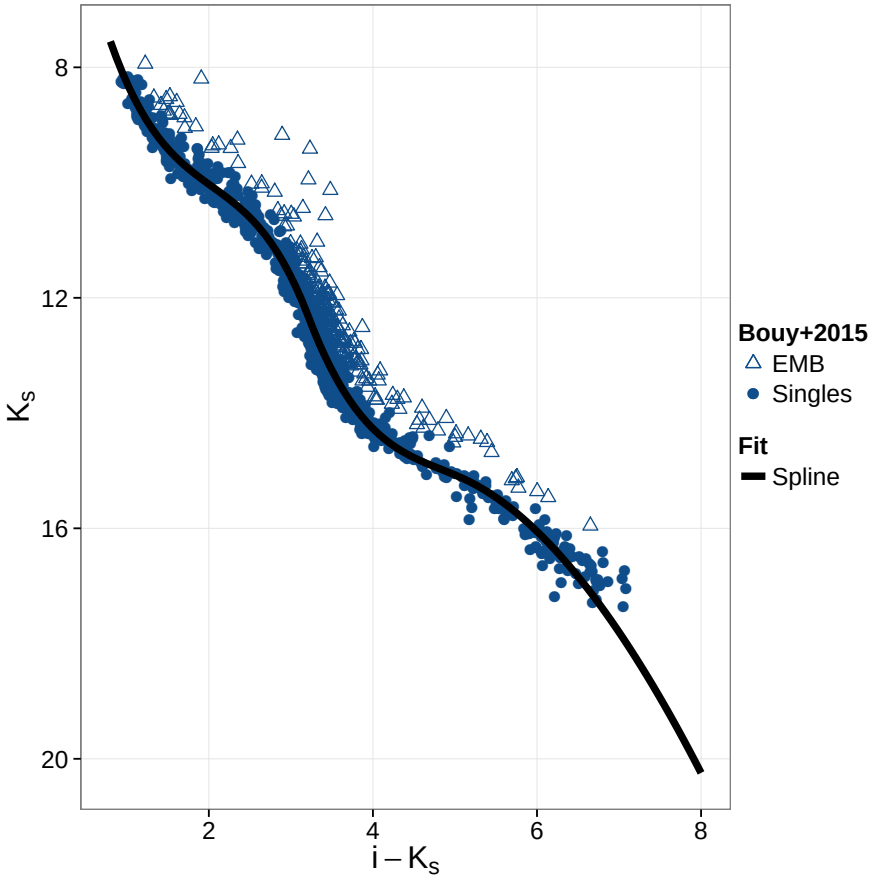
Assumptions

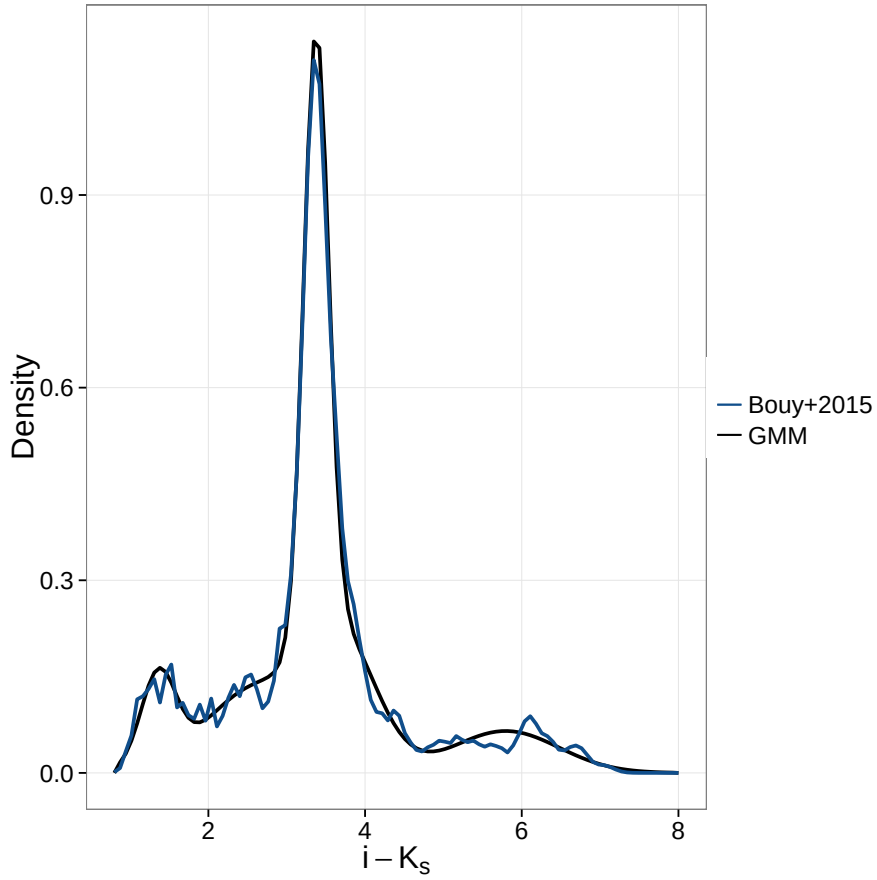
- Number of GMM components given by BIC
- Splines parameter: $i - K_s$
- Splines knots held fixed

Cluster proper motions model



Cluster photometric model





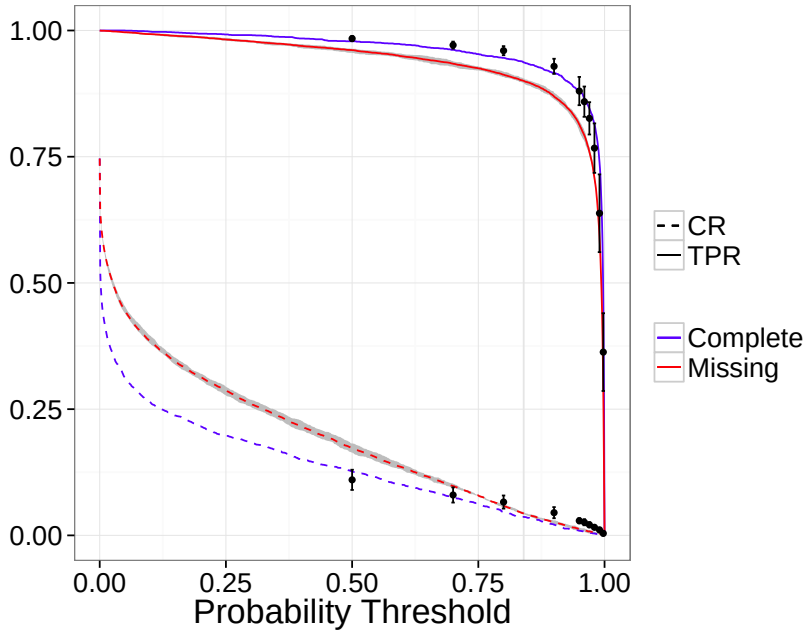
Obtaining the cluster distributions

High-Performance Computing

- Diverse infrastructures: UCA,UGA,CAB,UB
- Hybrid MPI-Multithreading approach
- Heuristic Particle Swarm Optimiser
- Markov Chain Monte Carlo
- Convergence

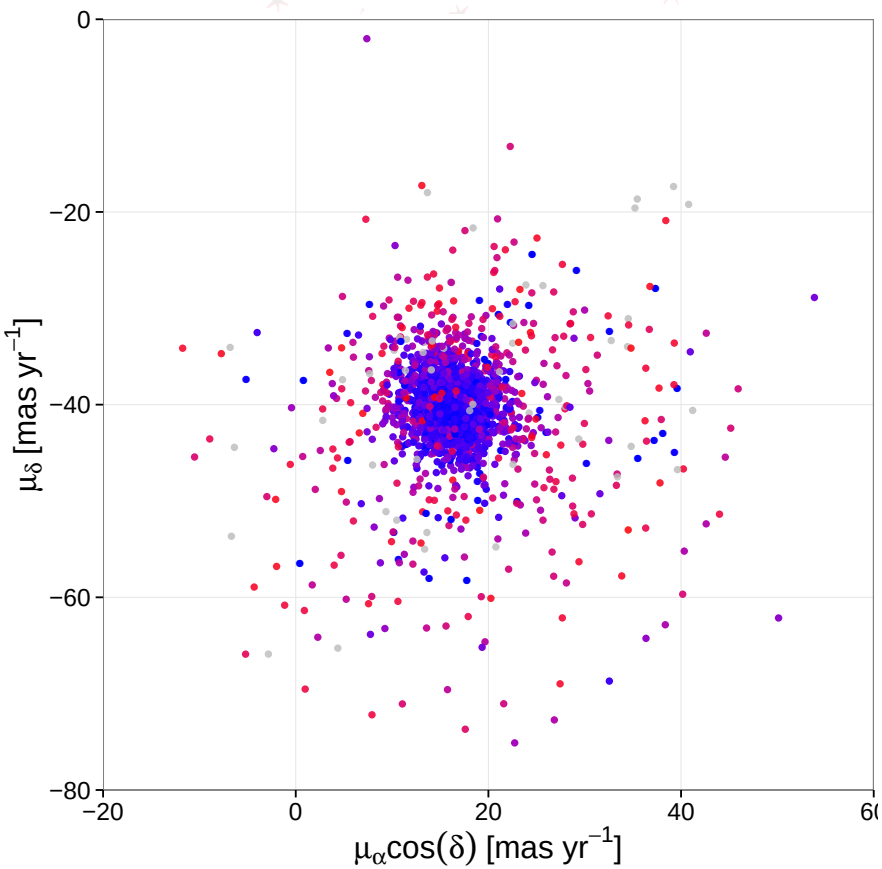
Validation on synthetic data

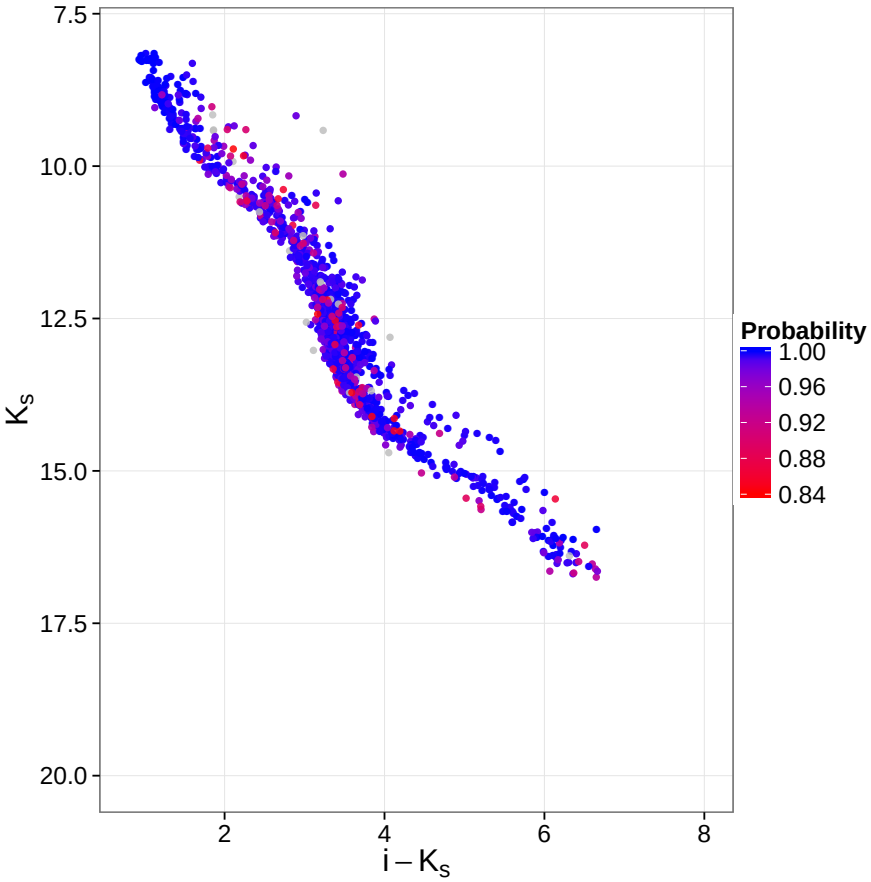
- Optimum Classification Threshold = 0.84
- Contamination Rate = $4.3 \pm 0.2\%$
- True Positive Rate = $90.0 \pm 0.05\%$
- Expected value of Contamination Rate = $5.8 \pm 0.2\%$



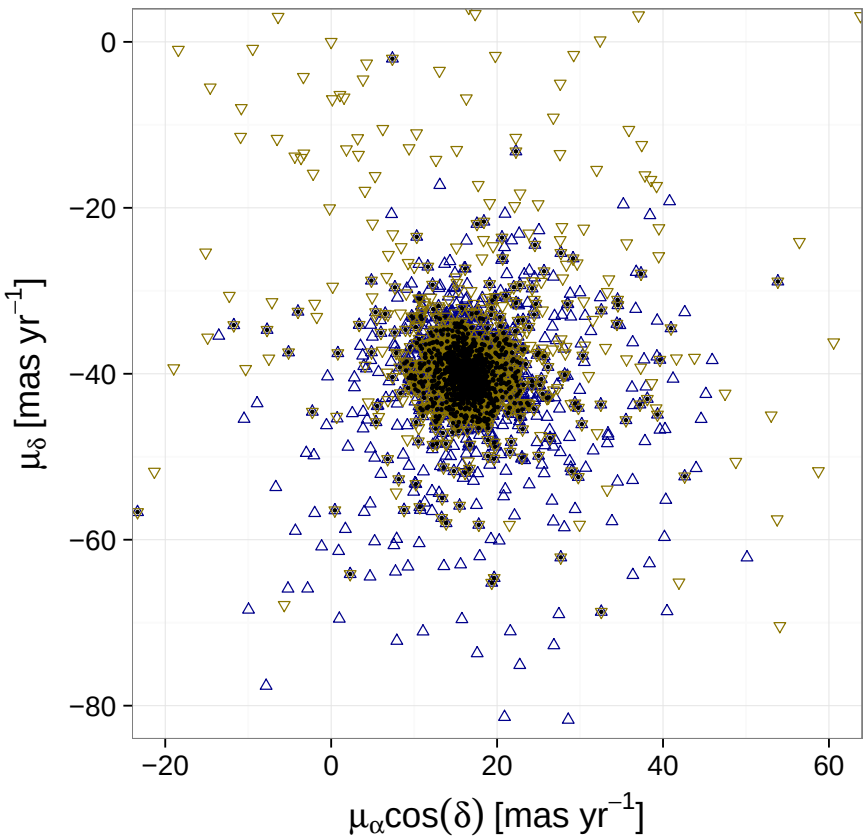
Results on the DANCe DR2

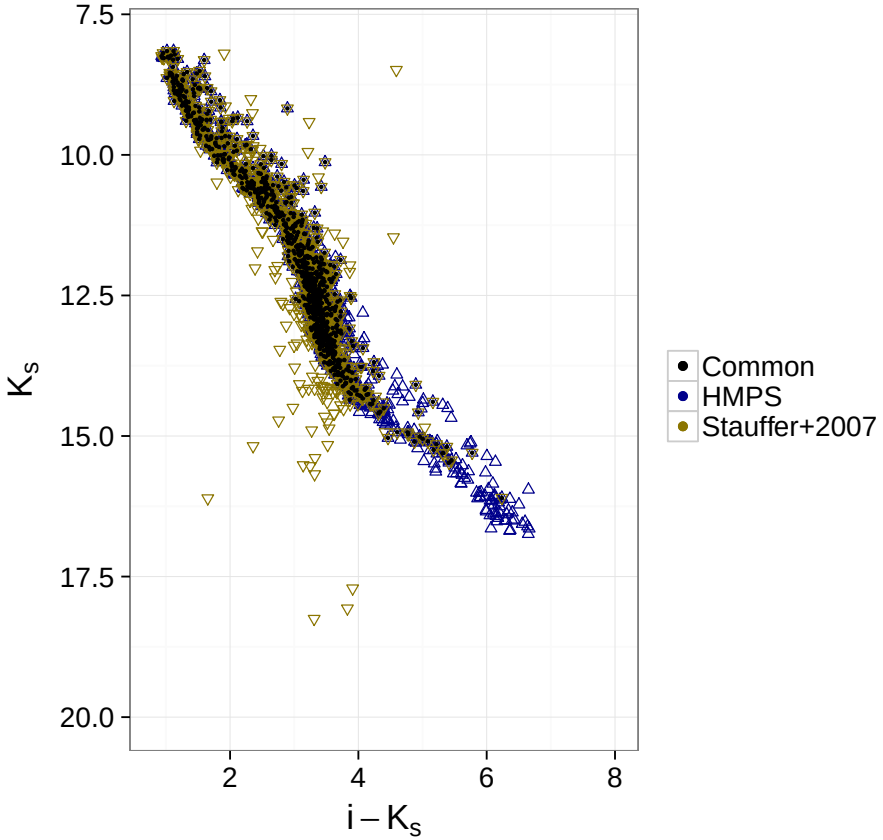
Cluster membership probabilities



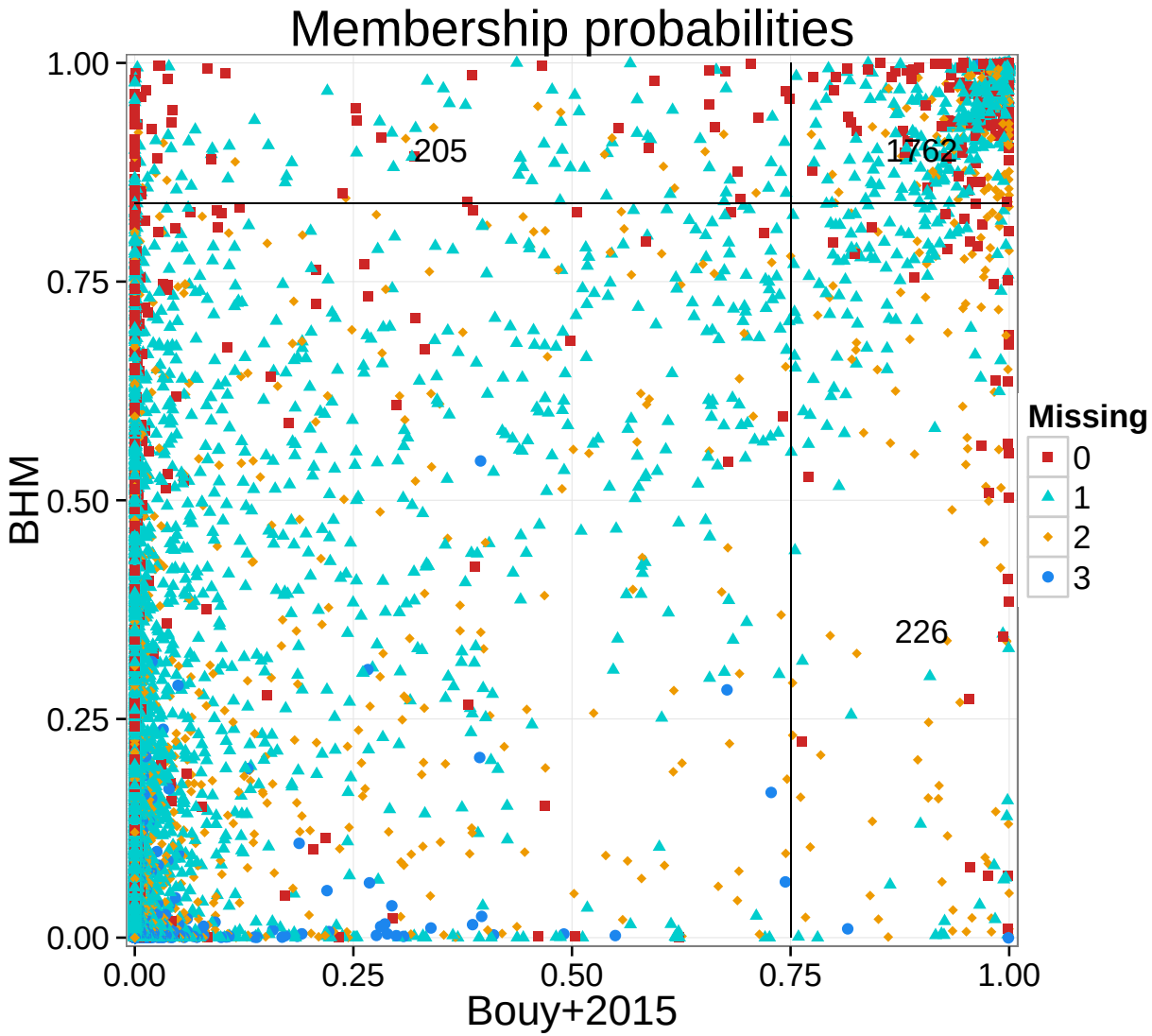


Evaluation: Stauffer et al. 2007

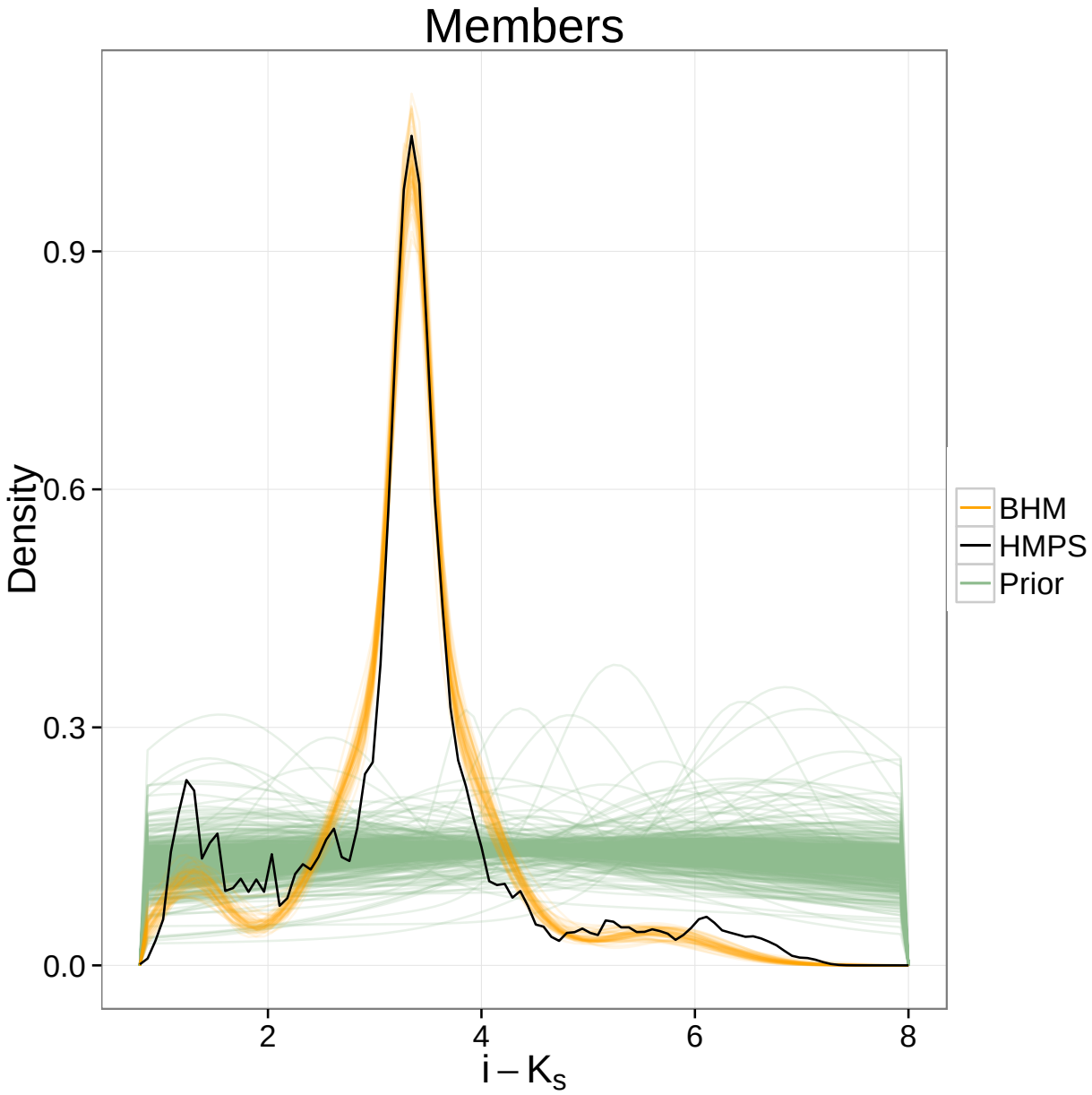




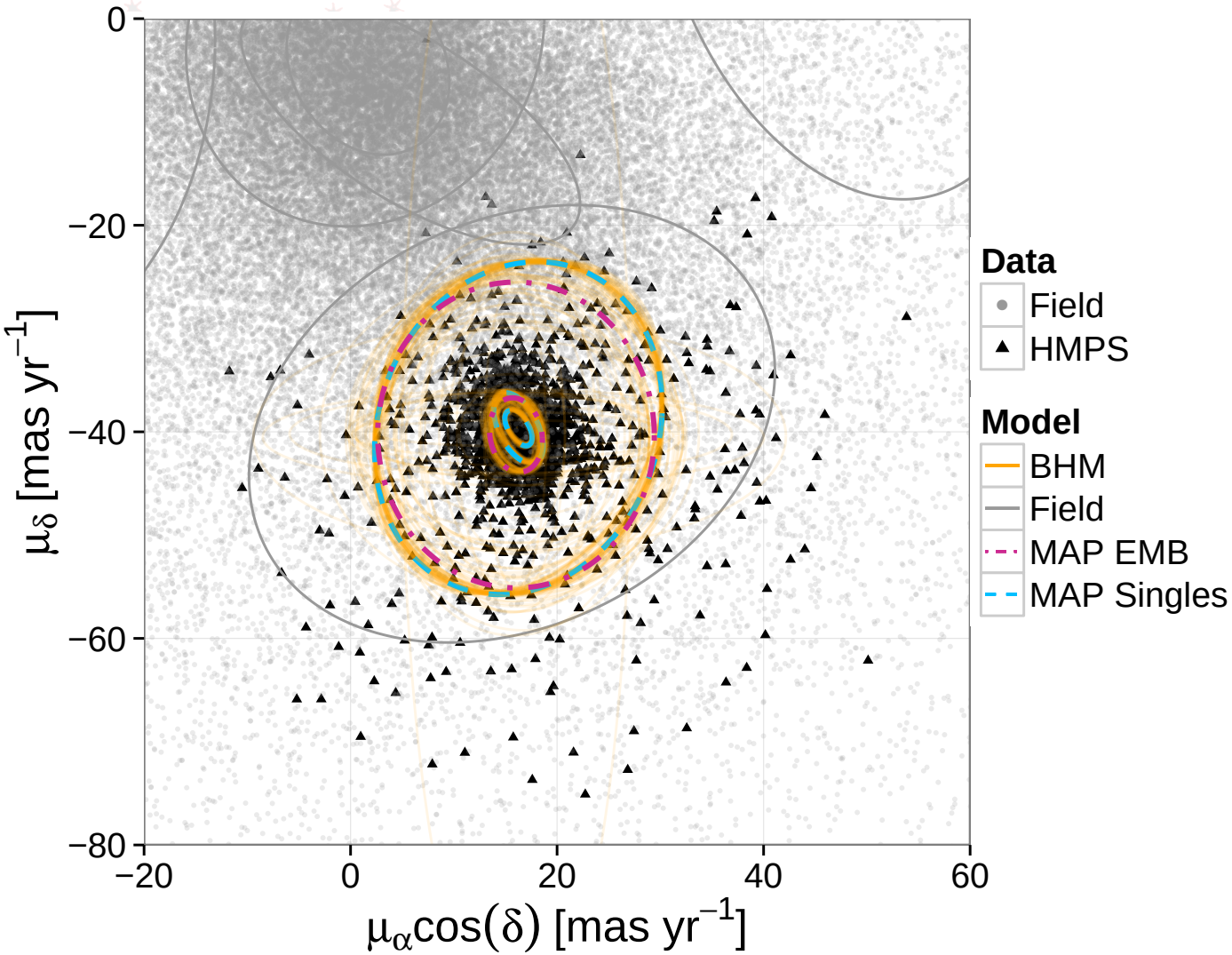
Evaluation: Bouy et al. 2015



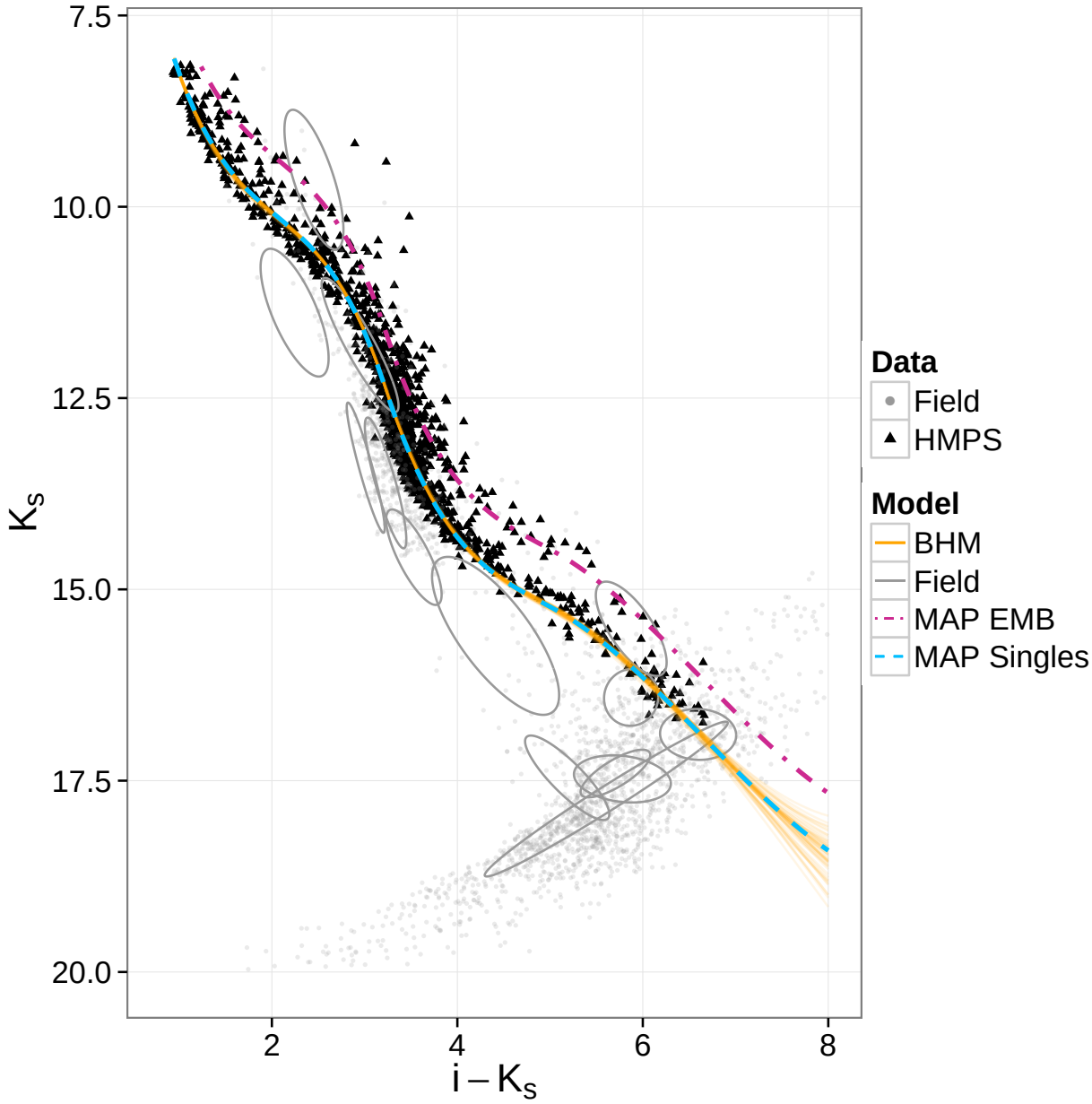
The statistical distributions: Colour index



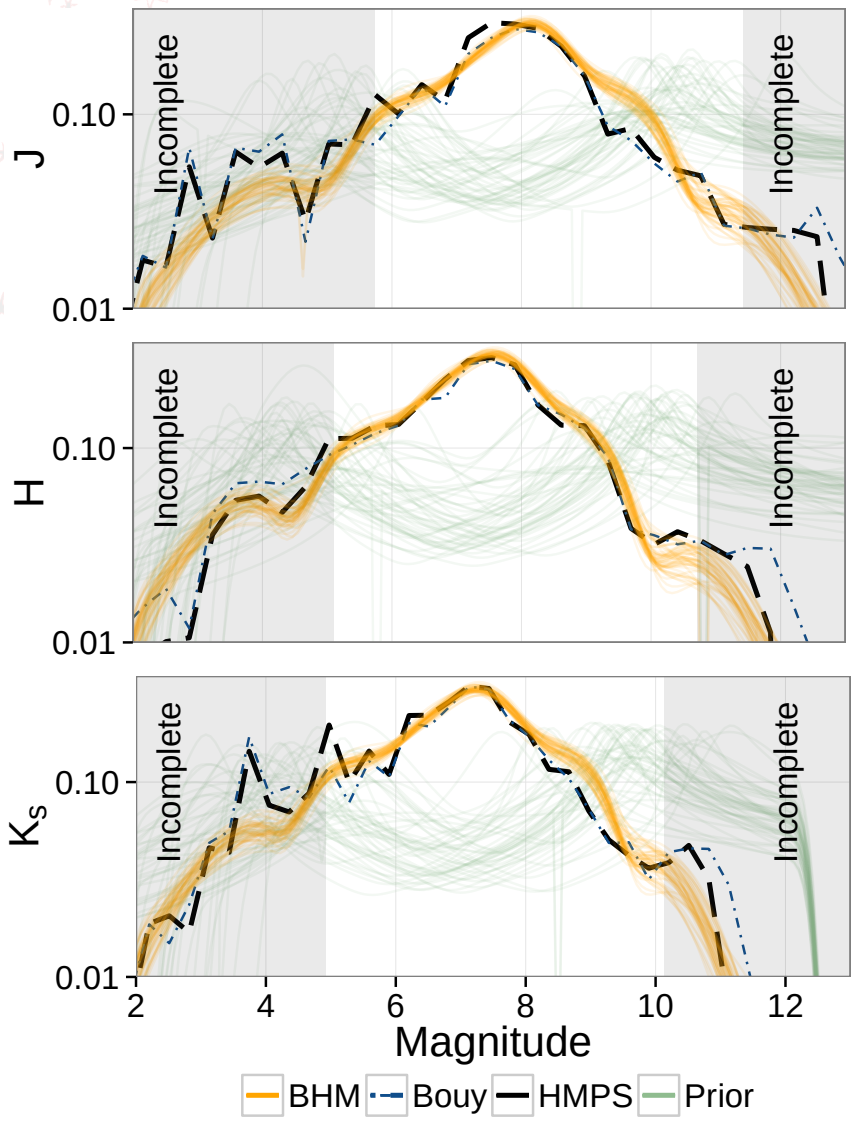
The statistical distributions: Proper motions



The statistical distributions: Empirical isochrone



The statistical distributions: Luminosity



The statistical distributions: Projected Spatial Distribution

Bayesian Model Selection

Models

- King 1962
- Elson et al. 1987
- Lauer et al. 1995
- Generalised King

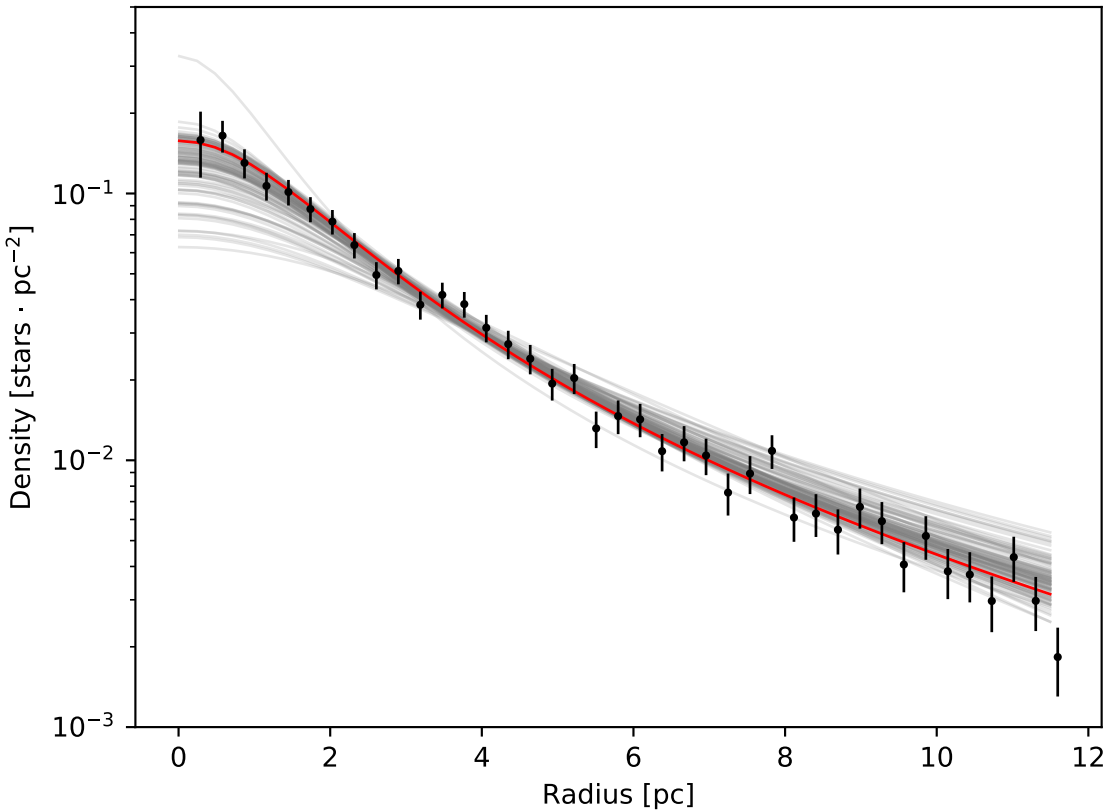
Extensions:

- Radially symmetric
- Biaxially symmetric
- Luminosity segregated

Simultaneously

Bayesian evidence + Parametric inference

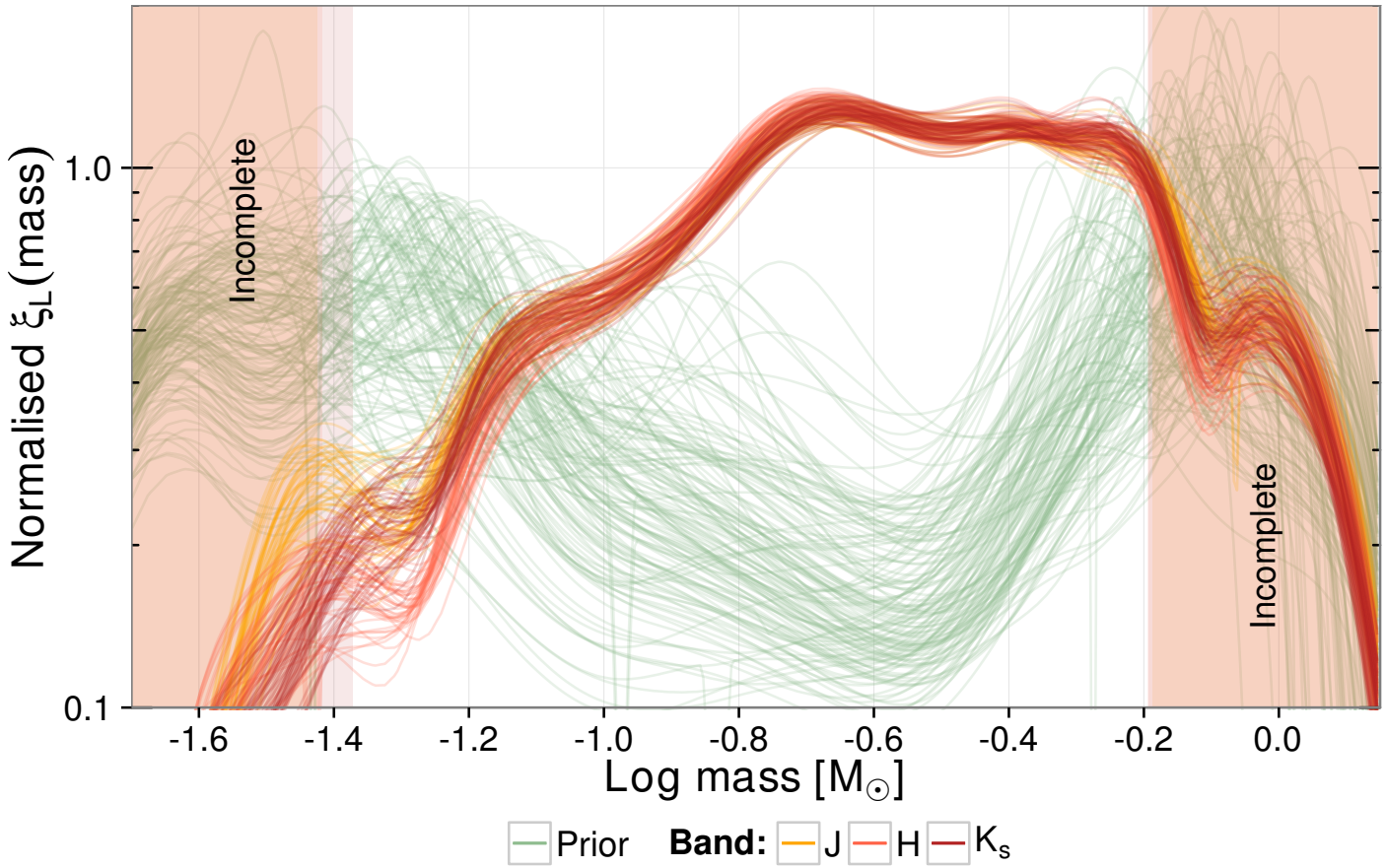
The statistical distributions: Projected Spatial Distribution



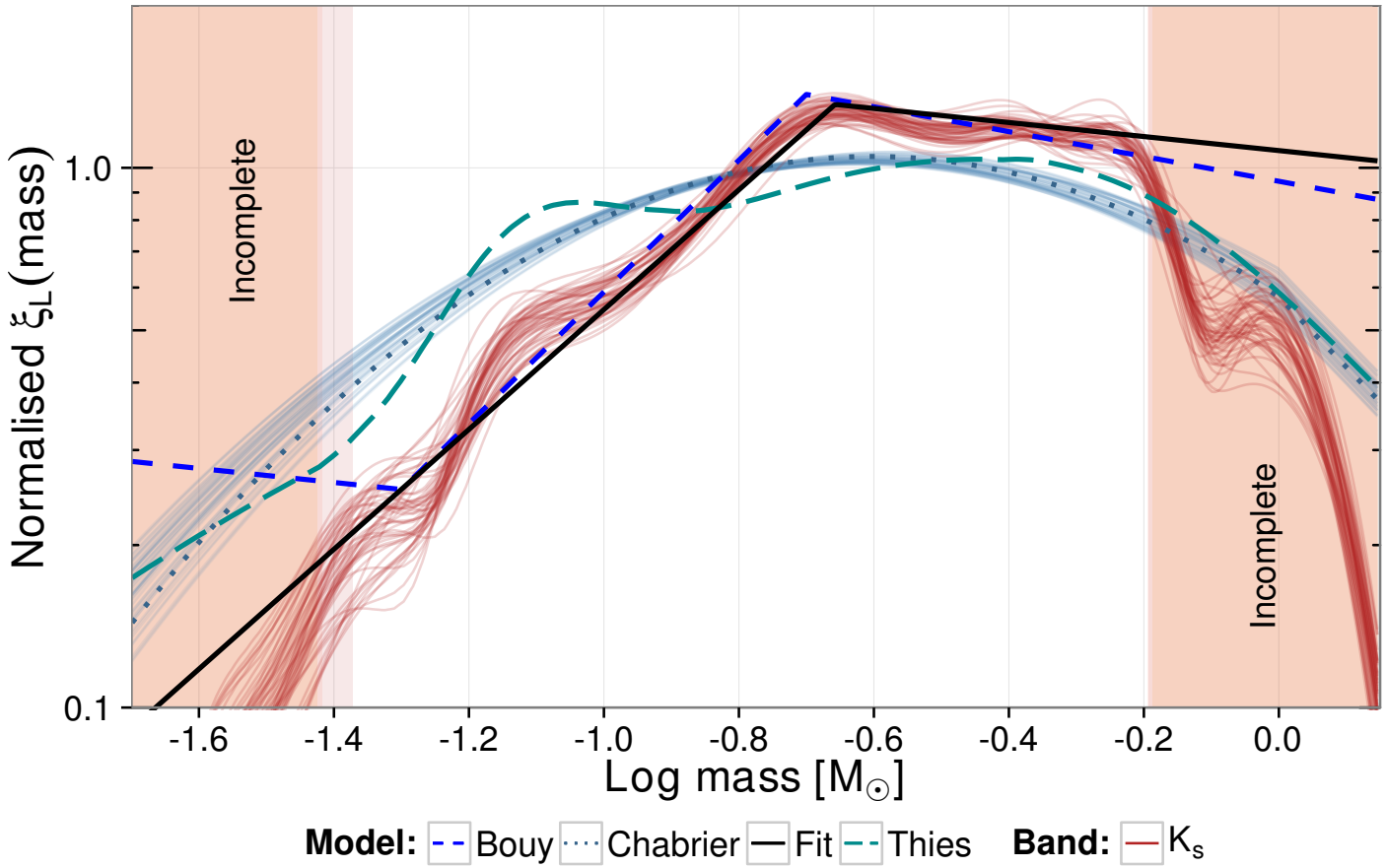
Results

- Strong evidence:
  - Ellipticity
  - Luminosity segregation
- Best models
  1. Generalised King
  2. King

Results: The mass distribution



Results: The mass distribution



Summary and conclusions

- Statistical distributions of the cluster population
  1. Constructed from 10^5 objects (missing values included)
  2. Non-homogeneous uncertainties propagated from all observables
- Membership probabilities for 2×10^6 objects
- 205 new cluster candidate members
- Discrepancy with IMF models low-mass domain
- Projected Spatial Distribution
  - Ellipticity
  - Mass segregation

Limitations and perspectives

- Field model
  - Include uncertainties
  - Infer it simultaneously with the cluster
- Missing values
  - Not at random!
- Computing time
- Parallax, Radial velocities, Extinction, Binaries

Limitations and perspectives

- Data from: Pan-STARRS, Gaia, LSST, etc.
- Liberate the software
- Use empirical isochrones to improve stellar models.
- COSMIC-DANCe: Compilation of luminosity distributions from Nearby Young Clusters